# On the Tightness of an LP Relaxation for Rational Optimization and Its Applications

Vashist Avadhanula, Jalaj Bhandari, Vineet Goyal and Assaf Zeevi[a]

[a]*Columbia University, New York, NY*

## Abstract

We consider the problem of optimizing a linear rational function subject to totally unimodular (TU) constraints over {0, 1} variables. Such formulations arise in many applications including assortment optimization. We show that a natural extended LP relaxation of the problem is "tight". In other words, any extreme point corresponds to an integral solution. We also consider more general constraints that are not TU but obtained by adding an arbitrary constraint to the set of TU constraints. Using structural insights about extreme points, we present a polynomial time approximation scheme (PTAS) for the general problem.

*Keywords:* Linear Programming, Assortment Optimization, PTAS

## 1. Introduction

We consider the following problem of optimizing a linear rational function over {0, 1} variables subject to totally unimodular (TU) constraints:

$$
\begin{aligned}
\text{maximize} \quad & \frac{a_0 + \sum_{i=1}^n a_i x_i}{c_0 + \sum_{j=1}^n c_j x_j} \\
\text{subject to} \quad & \mathbf{A}\mathbf{x} \le \mathbf{b} \\
& \mathbf{x} \in \{0, 1\}^n,
\end{aligned}
\tag{1}
$$

where $\mathbf{A}$ is a TU matrix, $\mathbf{b} \in Z^m$, $c_i \ge 0$ for all $i$. The above rational optimization problem arises in many applications including combinatorial optimization problems such as minimum mean cycle, minimum ratio shortest path, and assortment optimization over a Multinomial logit (MNL) choice model.

The above problem has been widely studied in the literature. In a seminal paper, Megiddo (1979) considers the problem of optimizing a rational objective over a combinatorial set, $D$ and presents a strongly polynomial algorithm if there is a combinatorial algorithm (only involving addition, subtraction and comparison operations) to optimize a linear objective over the set $D$. Hashizume et al. (1987), Correa et al. (2010) extend the work of Megiddo (1979) to provide approximation algorithms for the rational optimization problem over a combinatorial set assuming there is an approximation algorithm for linear optimization over the combinatorial

set. Mittal and Schulz (2013) provides a fully polynomial time approximation scheme (FPTAS) for optimizing a rational objective over a polytope and their technique can be extended to the case of a combinatorial set, if there is an efficient separation algorithm over the convex hull. Davis et al. (2014) consider a version of the rational optimization problem (1) in the context of assortment optimization over the MNL choice model and give an efficient algorithm to compute the optimal solution. Therefore, optimal algorithms for several close variants of (1) are known in the literature.

Our goal in this paper is to study the properties of a natural LP relaxation for the rational optimization problem (1). In particular, we study the tightness of the LP relaxation for the case of totally unimodular (TU) constraints. One advantage of the LP based approach for (1) is that it enables us to design approximation schemes for more general constraints; for instance, those obtained by adding an arbitrary constraint to the set of TU constraints.

### 1.1. Our Contributions.

Our main contributions are the following.

**Tight LP Relaxation**. We show that a natural LP relaxation of the rational optimization problem (1) under TU constraints is tight. We would like to note that, Davis et al. (2014) also use the LP relaxation to show that the assortment optimization under TU constraints can be solved optimally under the MNL choice model.

However, they do not explicitly analyze the structure of extreme points of the LP relaxation. In this paper, we show that an extreme point optimal solution for the LP relaxation is "integral" and therefore, gives an optimal solution for the rational optimization problem (1) under TU constraints. This structural property of the extreme points of the LP relaxation allows us to obtain near-optimal solutions for more general set of constraints that we discuss below.

**Extensions to More General Constraints**. We also consider the rational optimization problem (1) under more general constraints that are not necessarily TU. In particular, we consider an arbitrary additional constraint to the set of TU constraints such that the resulting set of constraints are not TU. We present a Polynomial Time Approximation Scheme (PTAS) for the rational optimization problem (1) under this more general set of constraints where for any $0 < \epsilon < 1$, we obtain a solution with objective value at least $(1-\epsilon)$ times the optimal in running time polynomial in the input size for a fixed $\epsilon$.

Our PTAS is based on the structure of the extreme points of the LP relaxation of (1) under this more general class of constraints. In particular, we show that any extreme point of the LP relaxation of (1) has only a small number of fractional components. After appropriate pruning, we can ignore the fractional components and obtain a feasible solution with near-optimal objective value.

We present applications of our PTAS approach to a class of constrained assortment optimization problem, namely, joint assortment and display optimization with capacity constraint. In this assortment optimization problem, we model the effect of display slot on the attraction parameter for any product and the goal is to select a subset of items and their display slots such that the capacity constrained is satisfied and the expected revenue is maximized. For the Multinomial logit (MNL) choice model, this can be modeled as a rational optimization problem (1) over a set of TU constraints and an additional capacity constraint. This problem is known to be NP-hard Desir and Goyal (2014). To the best of our knowledge, this is the first near-optimal approximation algorithm for the capacity constrained joint assortment and display problem. Desir and Goyal (2014) give an FPTAS for the assortment optimization under a capacity constraint for the MNL choice model. However, that approach does not extend to handle display constraints.

**Outline.** The rest of the paper is organized as follows.

In Section 2, we prove tightness of the LP relaxation. In Section 3, we present applications of the rational optimization problem (1) including cardinality constrained assortment optimization and joint assortment and display optimization with cardinality constraints. In Section 4, we present extensions to more general constraint sets and applications thereof.

**Notation.** We use the following notations in this paper. We use bold font to denote all vectors and matrices. For any matrix $\mathbf{Y} \in R^{m \times n}$ and index set $T \subset \{1, 2, \ldots, m\}$, $\mathbf{Y}(T)$ denotes the sub-matrix corresponding to rows $T$. The identity matrix is denoted as $\mathbf{I}$ and $\mathbf{e}$ denotes the vector of all ones of appropriate dimension. All vectors are column vectors. For any $n \in \mathcal{N}$, $[n]$ denotes the set $\{1, 2, \ldots, n\}$.

## 2. Rational Optimization: LP relaxation

In this section, we present a LP relaxation for (1) and show that the formulation is tight. Let

$$p_0 = \frac{1}{c_0 + \sum_{j=1}^n c_j x_j}, \quad p_i = x_i p_0, \ \forall i = 1, \ldots, n.$$

We can reformulate (1) as follows.

$$
\begin{aligned}
\underset{(\mathbf{p}, p_0)}{\text{maximize}} \quad & \sum_{i=0}^n a_i p_i \\
\text{subject to} \quad & \mathbf{A}\mathbf{p} \le p_0 \mathbf{b} \\
& \sum_{i=0}^n c_i p_i = 1 \\
& p_i \in \{0, p_0\} \ \forall \ i \in \{1, 2, \cdots, n\} \\
& p_0 \ge 0.
\end{aligned}
\tag{2}
$$

Note that (2) is an exact reformulation of (1). It can be easily reformulated as a mixed integer program using binary variables as follows.

$$
\begin{aligned}
& 0 \le p_i \le x_i p_0 \ \forall \ i \in \{1, 2, \cdots, n\} \\
& p_i + (1 - x_i) \ge p_0 \ \forall \ i \in \{1, 2, \cdots, n\} \\
& x_i \in \{0, 1\} \ \forall \ i \in \{1, 2, \cdots, n\}.
\end{aligned}
\tag{3}
$$

2

### 2.1. Tightness of the LP relaxation

We consider the following LP relaxation for (2).

$$z_{\mathsf{LP}} = \max_{(\mathbf{p}, p_0)} \sum_{i=1}^{n} a_i p_i$$

$$\mathbf{Ap} \le p_0 \mathbf{b}$$

$$\sum_{j=0}^{n} c_j p_j = 1 \tag{4}$$

$$0 \le p_i \le p_0, \ \forall i = 1, \dots, n,$$

$$p_0 \ge 0.$$

where we relax the constraints $p_i \in \{0, p_0\}$ to $0 \le p_i \le p_0$ for all $i = 1, \dots, n$. Let $\mathcal{P}$ be the polytope defined by the constraints in (4), i.e.,

$$\mathcal{P} = \left\{ (\mathbf{p}, p_0) \in \mathbb{R}_+^n \times \mathbb{R}_+ \mid \mathbf{Ap} \le p_0 \mathbf{b}, \ \mathbf{c'p} + c_0 p_0 = 1, \right.$$

$$\left. 0 \le p_i \le p_0, \forall i \right\}. \tag{5}$$

We show that all extreme points of $\mathcal{P}$ are "integral". We say that an extreme point $(\mathbf{p}, p_0) \in \mathcal{P}$ is *integral* if $p_i \in \{0, 1\}$ for all $i = 1, \dots, n$ and *fractional* otherwise.

**Theorem 1.** *For any extreme point* $(\mathbf{p}, p_0) \in \mathcal{P}$, $p_i \in \{0, p_0\}$ *for all* $i = 1, \dots, n$.

We will prove Theorem 1 by establishing a correspondence between extreme points of $\mathcal{P}$ and $Q$, where

$$Q = \{\mathbf{x} \mid \mathbf{Ax} \le \mathbf{b}, \ 0 \le x_i \le 1 \ \text{for all} \ i = 1, 2, \dots, n\},$$

is the polytope corresponding to relaxed constraints of the rational optimization problem (1).

**Lemma 1.** *If* $(\mathbf{p}, p_0)$ *is an extreme point of* $\mathcal{P}$, *then* $\mathbf{x} = \frac{\mathbf{p}}{p_0}$ *is an extreme point of* $Q$. *Conversely, if* $\mathbf{x}$ *is an extreme point of* $Q$, *then* $(\mathbf{p}, p_0)$ *where*

$$p_0 = \frac{1}{(c_0 + \mathbf{c'x})}, \ \mathbf{p} = p_0 \mathbf{x}$$

*is an extreme point of* $\mathcal{P}$.

*Proof.* Consider any extreme point, $(\mathbf{p}, p_0)$ of $\mathcal{P}$. Note that there must be $n + 1$ linearly independent and active constraints. Let

$$S_0 = \{i \mid p_i = 0\}, \ S_1 = \{i \mid p_i = p_0\},$$

$$T = \{i \mid \sum_{j=1}^{n} a_{ij} p_j = b_i p_0\} \tag{6}$$

$$k = |S_0| + |S_1| + |T|.$$

We claim that $k \ge n$. This follows as we have $|S_0| + |S_1|$ linearly independent and active constraints from the constraint set $S_0 \cup S_1$, $|T|$ active constraints from the constraint set $T$ and one active constraint from the constraint $\sum_{i=0}^{n} c_i p_i = 1$. Hence the total number of linearly independent and active constraints at $(\mathbf{p}, p_0)$ is at most $k + 1$.

Without loss of generality we can assume that $k = n$; since $k > n$ implies that $|S_0| + |S_1| + |T| + 1 > n + 1$, making some constraints in $T$ redundant. Let

$$\mathbf{B}_p = \begin{bmatrix} \mathbf{A}(T) & -\mathbf{b}(T) \\ \mathbf{I}(S_0) & \mathbf{0} \\ \mathbf{I}(S_1) & -\mathbf{e} \\ \mathbf{c'} & c_0 \end{bmatrix}, \quad \mathbf{B}_x = \begin{bmatrix} \mathbf{A}(T) \\ \mathbf{I}(S_0) \\ \mathbf{I}(S_1) \end{bmatrix}, \quad \mathbf{b}_x = \begin{bmatrix} \mathbf{b}(T) \\ \mathbf{0} \\ \mathbf{e} \end{bmatrix}, \tag{7}$$

Note that $\mathbf{B}_p$ is the basis matrix corresponding to the extreme point $(\mathbf{p}, p_0)$. Hence, $\mathbf{B}_p$ is full rank. For the sake of contradiction, assume that $\mathbf{B}_x$ is not full rank. There there exists $\boldsymbol{\lambda} \in R^n, \boldsymbol{\lambda} \ne \mathbf{0}$ such that $\boldsymbol{\lambda'} \mathbf{B}_x = \mathbf{0}$, then we have

$$\begin{bmatrix} \boldsymbol{\lambda'} & 0 \end{bmatrix} \mathbf{B}_p = \begin{bmatrix} \boldsymbol{\lambda'} \mathbf{B}_x & -\boldsymbol{\lambda'} \mathbf{b}_x \end{bmatrix} = \begin{bmatrix} \mathbf{0} & -\boldsymbol{\lambda'} \mathbf{b}_x \end{bmatrix},$$

which implies

$$\begin{bmatrix} \boldsymbol{\lambda'} & 0 \end{bmatrix} \mathbf{B}_p \begin{bmatrix} \mathbf{p} \\ p_0 \end{bmatrix} = -p_0 \boldsymbol{\lambda'} \mathbf{b}_x,$$

Since $\mathbf{B}_p$ is a full rank, we have $\boldsymbol{\lambda'} \mathbf{b}_x \ne 0$ and $p_0 \ne 0$, contradicting that,

$$\mathbf{B}_p \begin{bmatrix} \mathbf{p} \\ p_0 \end{bmatrix} = \begin{bmatrix} \mathbf{0} \\ 1 \end{bmatrix}.$$

Hence, $\mathbf{B}_x$ is a full rank and $\mathbf{x} = \mathbf{p}/p_0$ is a basic feasible solution in $Q$ corresponding to the basis matrix $\mathbf{B}_x$.

Conversely, consider $\mathbf{x}$, any extreme point of $Q$. Let

$$p_0 = \frac{1}{c_0 + \mathbf{c'x}}, \quad \mathbf{p} = p_0 \mathbf{x}.$$

Clearly $(\mathbf{p}, p_0) \in \mathcal{P}$. We define the quantities $S_0, S_1, T, k$ as in (6) and $\mathbf{B}_p, \mathbf{B}_x, \mathbf{b}_x$ as in (7). Using similar arguments, we can assume $k = n$ without loss of generality. Since $\mathbf{x}$ is a basic feasible solution corresponding to the basis $\mathbf{B}_x$, $\mathbf{B}_x$ is full rank.

For the sake of contradiction, suppose $\mathbf{B}_p$ is not full rank. Then there exists $\boldsymbol{\lambda} \in R^{n+1}, \boldsymbol{\lambda} \ne \mathbf{0}$ such that $\boldsymbol{\lambda'} \mathbf{B}_p = \mathbf{0}$. Therefore,

$$\boldsymbol{\lambda'} \mathbf{B}_p \begin{bmatrix} \mathbf{p} \\ p_0 \end{bmatrix} = 0$$

which implies

$$(\boldsymbol{\lambda}([n]))'(\mathbf{B}_x \mathbf{p} + p_0 \mathbf{b}_x) + \lambda_{n+1}(\mathbf{c'p} + c_0 p_0) = 0.$$

Since $\mathbf{B}_x\mathbf{x} = \mathbf{b}_x$, we have $\mathbf{B}_x\mathbf{p} + p_0\mathbf{b}_x = \mathbf{0}$ and $\lambda_{n+1} = 0$. Note that, $\boldsymbol{\lambda}'\mathbf{B}_p = \mathbf{0}$ and

$$\boldsymbol{\lambda}'\mathbf{B}_p = \begin{bmatrix} \boldsymbol{\lambda}([n])'\mathbf{B}_x + \lambda_{n+1}\mathbf{c}' & \boldsymbol{\lambda}([n])'\mathbf{b}_x + \lambda_{n+1}c_0 \end{bmatrix}$$

Therefore $\boldsymbol{\lambda}([n])'\mathbf{B}_x = \mathbf{0}$, contradicting the fact that $\mathbf{B}_x$ is full rank. Hence, $\mathbf{B}_p$ is a full rank matrix and $(\mathbf{p}, p_0)$ is the basic feasible solution corresponding to the basis matrix $\mathbf{B}_p$. This completes the proof. $\square$

Theorem 1 follows from Lemma 1 and the fact that any extreme point $\mathbf{x}$ of $Q$ is integral, i.e. $x_i \in \{0, 1\}$. We would like to emphasize that Davis et al. (2014) take a similar approach of reformulating the rational optimization problem (1) as LP relaxation (4). However, they do not explicitly analyze the structure of extreme points of (4). Theorem 1 proves that any extreme point optimal solution of LP relaxation (4) is the same as the MIP reformulation (2) and hence it suffices to solve the relaxation.

## 3. Applications: Assortment Optimization

Assortment optimization problems arise in many applications including retailing where a seller needs to select a subset of substitutable products to offer to consumers exhibiting substitution behavior such that the expected revenue is maximized. If the substitution behavior is given by a Multinomial logit (MNL) choice model, the assortment optimization problem can be formulated as a special case of (1). Davis et al. (2014) consider the assortment optimization problem over the MNL choice model under totally-unimodular constraints and present several applications. In order to illustrate the usefulness of our LP approach for more general constraints than TU, we describe two special cases below: cardinality constrained assortment optimization and joint assortment and display optimization for the MNL choice model.

**Cardinality Constrained Assortment Optimization**: In a cardinality constrained assortment optimization problem, there is an upper bound on the total number of products offered in the assortment. For the MNL choice model, this problem can be formulated as a rational objective over TU constraints. Let $n$ be the total number of products and $K$ be the bound on total number of products that can be offered. For any $i \in [n]$, let $x_i$ be a binary variable that denotes whether product $i$ is offered or not. Also, let $v_i$ denote the attraction parameter (or mean utility) and $r_i$ is the revenue of product $i$. Then

the problem of maximizing the expected revenue can be formulated as:

$$\underset{\mathbf{x}\in\{0,1\}^n}{\text{maximize}} \quad R(\mathbf{x}) = \frac{\sum_{i=1}^n r_i v_i x_i}{v_0 + \sum_{i=1}^n v_i x_i}$$
$$\text{subject to} \quad \sum_{i=1}^n x_i \le K \tag{8}$$
$$x_i \in \{0, 1\} \ \forall \ i.$$

**Joint Assortment and Display Optimization**: In a joint assortment and display optimization problem, the retailer needs to select the subset of products to offer and also decide on the display segment. This problem arises in retailing and online advertising where the display slot of the product/ad affects the choice probability. In particular, we consider a model with $m$ display segments and each segment has an upper bound on the number of products that can be displayed. The customers choose according to an MNL model, where the purchasing probability of each offered product also depends on its display segment. The objective is to compute an optimal assortment together with the optimal display segment for each offered product such that the cardinality constraints for each segment are satisfied and the expected revenue is maximized.

Let $n$ be the total number of products and $m$ be the number of display segments. There is a bound $K_j$ on the number of products in display segment $j$ for all $j \in [m]$. We assume that every product can only be displayed in at most one display segments. Let $x_{ij} \in \{0, 1\}$ denote whether we offer product $i$ in display segment $j$. For any product $i$, let $r_i$ denote the revenue and $v_{ij}$ denote the attraction parameter in display segment $j$. Now, the expected revenue optimization problem can be formulated as:

$$\underset{\mathbf{X}\in\{0,1\}^{n\times m}}{\text{maximize}} \quad R(\mathbf{X}) = \frac{\sum_{i=1}^n \sum_{j=1}^m r_i v_{ij} x_{ij}}{v_0 + \sum_{i=1}^n \sum_{j=1}^m \hat{v}_{ij} x_{ij}}$$

$$\text{subject to} \quad \mathbf{C_i} : \quad \sum_{j=1}^m x_{ij} \le 1, \ i = 1, \dots, n$$

$$\mathbf{C_j} : \quad \sum_{i=1}^n x_{ij} \le K_j, \ j = 1, \dots, m$$

$$x_{ij} \in \{0, 1\}, \ i = 1, \dots, n, j = 1, \dots, m. \tag{9}$$

Constraints $\{\mathbf{C_i}\}$ enforce that every product can be displayed only in one of the display segments, while constraints $\{\mathbf{C_j}\}$ enforce the cardinality constraints in each segment. The constraints in problem (9) are identical to the constraints in a transportation problem and hence

4

are TU, making problem (9) a special case of the rational optimization problem (1).

## 4. Extension to More General Constraints

In this section, we consider a more general variant of the rational optimization problem (1), where constraints are not necessarily TU. In particular, we consider the following problem where we have a set of TU constraints and one additional constraint such that the overall constraints are not TU:

$$
\begin{aligned}
\text{maximize} \quad & \frac{a_0 + \sum_{i=1}^{n} a_i x_i}{c_0 + \sum_{j=1}^{n} c_j x_j} \\
\text{subject to} \quad & \mathbf{A}\mathbf{x} \le \mathbf{b} \\
& \boldsymbol{\alpha}^T \mathbf{x} \le \gamma \\
& \mathbf{x} \in \{0, 1\}^n,
\end{aligned}
\tag{10}
$$

where $\mathbf{A}$ is a $\{0, 1\}^{m \times n}$ TU matrix, $\mathbf{b} \in Z^m$, $c_i \ge 0$ and $\alpha_i \ge 0$ for all $i$. Let

$$
Q = \{\mathbf{x} \mid A\mathbf{x} \le \mathbf{b},\ 0 \le x_i \le 1 \text{ for all } i = 1, 2, \ldots, n\}
$$
$$
\hat{Q} = \left\{ \mathbf{x} \in Q \mid \boldsymbol{\alpha}^T \mathbf{x} \le \gamma \right\},
$$

be the polytopes corresponding to the relaxations of (1) and (10) respectively.

Similar to our approach in Section 2, we consider the following LP relaxation for (10),

$$
\begin{aligned}
\underset{(\mathbf{p}, p_0)}{\text{maximize}} \quad & \sum_{i=0}^{n} a_i p_i \\
\text{subject to} \quad & (\mathbf{p}, p_0) \in \mathcal{P} \\
& \boldsymbol{\alpha}^T \mathbf{p} \le p_0 \gamma.
\end{aligned}
\tag{11}
$$

where $\mathcal{P}$ is the polytope corresponding to the LP relaxation of the rational optimization problem (1) as defined in (5). Let

$$
\hat{\mathcal{P}} = \left\{ (\mathbf{p}, p_0) \in \mathcal{P} \mid \boldsymbol{\alpha}^T \mathbf{p} \le p_0 \gamma \right\},
$$

be the polytope corresponding to the LP relaxation of (10). Since constraints in (10) are not TU, the LP relaxation (11) may not be tight. In this section, we present a polynomial time approximation scheme (PTAS) for (10) under certain assumptions on $Q$. In other words, for a fixed $\epsilon$, we compute a $(1 - \epsilon)$-approximation for (10) in running time polynomial in the input but exponential in $1/\epsilon$. Our PTAS is based on the following structure of extreme points of (11).

Observe that the polytope $\hat{Q}$ (respectively $\hat{\mathcal{P}}$) is the intersection of the polytope $Q$ (respectively $\mathcal{P}$) and the hyperplane $\boldsymbol{\alpha}^T \mathbf{x} \le \gamma$ (respectively $\boldsymbol{\alpha}^T \mathbf{p} \le p_0 \gamma$). Hence, any

extreme point of $\hat{Q}$ (respectively $\hat{\mathcal{P}}$) is either an extreme point of $Q$ (respectively $\mathcal{P}$) or a convex combination of two adjacent extreme points of $Q$ (respectively $\mathcal{P}$). Therefore, if any two adjacent extreme points of $Q$ "differ" only in a small number of components, then the number of "fractional components" in any extreme point of $\hat{Q}$ and $\hat{\mathcal{P}}$ is small. Therefore, we can obtain an approximate solution for (10) by ignoring the small number of "fractional components" from the optimal solution of (11) after appropriate pruning.

More specifically, for any two extreme points $\mathbf{x}_1, \mathbf{x}_2$ of $Q$, let

$$
d(\mathbf{x}_1, \mathbf{x}_2) = |\{i \mid x_{1i} \ne x_{2i}\}|
$$
$$
d(Q) = \max \{d(\mathbf{x}_1, \mathbf{x}_2) \mid \mathbf{x}_1, \mathbf{x}_2 \text{ are adjacent extr pts of } Q\}.
$$

Here $d(Q)$ denotes the maximum number of components by which the two adjacent extreme points of $Q$ can differ. If $d(Q) \le \ell$, then the number of fractional components for any extreme point of $\hat{Q}$ is at most $\ell$. From Lemma 1, we know that there is a correspondence between extreme points of $\mathcal{P}$ and $Q$. A similar correspondence also holds for extreme points of $\hat{\mathcal{P}}$ and $\hat{Q}$. Hence, the number of "fractional components" in any extreme point of $\hat{\mathcal{P}}$ is also bounded by $\ell$. In particular, for any extreme point $(\mathbf{p}, p_0)$ of $\hat{\mathcal{P}}$, let

$$
\mathcal{F}((\mathbf{p}, p_0)) = \{i \ge 1 \mid 0 < p_i < p_0\},
$$

denote the set of fractional components in $(\mathbf{p}, p_0)$. We have the following result,

**Corollary 1.** *If $d(Q) \le \ell$, then the number of fractional components for any extreme point $(\mathbf{p}, p_0)$ of $\hat{\mathcal{P}}$ is bounded by $\ell$, i.e. $|\mathcal{F}((\mathbf{p}, p_0))| \le \ell$.*

### 4.1. PTAS when $d(Q)$ is constant

In this section, we present a PTAS for the case when $d(Q)$ is a constant (say $\ell$). From Lemma 1, we know that optimality (feasibility) of $(\mathbf{p}, p_0)$ is equivalent to optimality (feasibility) of $\mathbf{x} = \mathbf{p}/p_0$ for (10). From Corollary 1, we know that any extreme point to (11) has at most $\ell$ fractional variables as $d(Q) = \ell$. A simple idea to construct a feasible solution for (10) from an optimal solution of (11) is to ignore the "fractional variables". In particular, let $(\mathbf{p}, p_0) \in \hat{\mathcal{P}}$ be an optimal extreme point of (11). Define $(\hat{\mathbf{p}}, \hat{p}_0)$ as

$$
\hat{p}_i = \begin{cases} 0 & \text{if } p_i < p_0 \\ \hat{p}_0 & \text{otherwise} \end{cases}
$$

where

$$
\hat{p}_0 = \frac{1}{c_0 + \sum_{i : \hat{p}_i \ne 0} c_i}.
$$

5

Observe that we ignore at most $\ell$ variables of $(\mathbf{p}, p_0)$. If the contribution of these variables to the objective value is small, then the total decrease in objective value is also bounded. Let $R^*$ denote the optimal objective value of (10). If

$$a_i p_i \leq \frac{\epsilon}{\ell} R^* \quad \forall\, i : \; p_i < p_0, \qquad (12)$$

then

$$\sum_{i \in \mathcal{F}(\mathbf{p}, p_0)} a_i p_i \leq \epsilon R^*,$$

which implies

$$(1 - \epsilon) R^* \leq \sum_{i \notin \mathcal{F}(\mathbf{p}, p_0)} a_i p_i = \sum_i a_i \hat{p}_i,$$

and $(\hat{\mathbf{p}}, \hat{p}_0)$ is a $(1 - \epsilon)$-approximate solution for (10). Note that in $(\mathbf{p}, p_0)$ there can be at most $\left\lfloor \frac{\ell}{\epsilon} \right\rfloor$ variables such that $a_i p_i > \frac{\epsilon}{\ell} R^*$. Therefore, to ensure (12) we guess the top $\left\lfloor \frac{\ell}{\epsilon} \right\rfloor$ variables contributing to the objective in (10), set those variables $p_i = p_0$ and solve the resulting linear program. The running time is exponential in $\left\lfloor \frac{\ell}{\epsilon} \right\rfloor$. Hence, we have the following result.

**Theorem 2.** *If $d(Q) \leq \ell$, then there exists an Algorithm that obtains an $(1 - \epsilon)$ approximate solution to (10) in running time that is polynomial in the input size for a fixed $\epsilon$.*

We present the PTAS and the proof of correctness of the algorithm in Appendix A.

### 4.2. Examples of $Q$ with small $d(Q)$

So far, we have assumed that $d(Q) \leq \ell$ and restricted our attention to specific instances of the rational optimization problem (1) that satisfy this criteria. There are large classes of problems that can be formulated in the framework of the rational optimization problem (1) and also satisfy the assumption that $d(Q)$ is small. In this section, we revisit applications discussed in Section 3 and establish that $d(Q)$ is indeed small, enabling our PTAS approach to solve the more generic version of these problems.

**Assortment Optimization with Cardinality Constraint:** The polytope $Q$ corresponding to the feasible region of cardinality constrained assortment optimization problem (8) is

$$Q = \left\{ \mathbf{x} \;\middle|\; \sum_{i=1}^n x_i \leq K, \; 0 \leq x_i \leq 1 \,,\; i = 1, 2, \ldots, n \right\}.$$

Note that the polytope $Q$ is the intersection of the n dimensional hypercube and the hyperplane $\sum_{i=1}^n x_i \leq K$. We know that every extreme point $\mathbf{x}$ of $Q$ is such that $\mathbf{x} \in \{0, 1\}^n$ and every pair of adjacent extreme points in the n dimensional hypercube only differ in two components. Hence, we have the following result

**Lemma 2.** *For $Q$ corresponding to the cardinality constrained assortment optimization problem (8), we have $d(Q) = 2$.*

**Joint Assortment and Display Optimization**: The polytope $Q$ corresponding to the feasible region of cardinality constrained joint assortment and display optimization problem (9) is

$$Q = \left\{ \mathbf{X} \;\middle|\; \sum_{j=1}^m x_{ij} \leq 1 \; \forall i, \; \sum_{i=1}^n x_{ij} \leq K_j \; \forall j, \; \mathbf{0} \leq \mathbf{X} \leq \mathbf{1} \right\}.$$

The constraints in problem (9) are the same as the transportation problem, the number of variables that are different in two adjacent extreme points of the LP relaxation of problem (9) is bounded by the maximum cycle length in the corresponding transportation network. Since the transportation network is a bipartite graph, the maximum cycle length cannot exceed twice the number of nodes in either of the partitions. Hence, we have the following result,

**Lemma 3.** *For $Q$ corresponding to feasible region of cardinality constrained joint assortment and display optimization problem (9), we have $d(Q) \leq 2m$, where m is the number of display segments.*

Theorem 3 and the Lemmas 2,3 establishes that there exists a PTAS for the above applications in the presence of an additional constraint.

### References

N. Megiddo, Combinatorial Optimization with Rational Objective Functions, Mathematics of Operations Research 4 (4) (1979) pp. 414–424.

S. Hashizume, M. Fukushima, N. Katoh, T. Ibaraki, Approximation algorithms for combinatorial fractional programming problems, Mathematical Programming 37 (3) (1987) 255–267, ISSN 0025-5610.

J. Correa, C. Fernandes, Y. Wakabayashi, Approximating a class of combinatorial problems with rational objective function, Mathematical Programming 124 (1-2) (2010) 255–269, ISSN 0025-5610.

S. Mittal, A. Schulz, An FPTAS for optimizing a class of low-rank functions over a polytope, Mathematical Programming 141 (1-2) (2013) 103–120, ISSN 0025-5610.

J. Davis, Gallego, G., H. Topaloglu, Assortment Planning under the Multinomial logit Model with Totally Unimodular Constraint Structures, Working Paper 14 (2014) 327–344.

A. Desir, V. Goyal, An FPTAS for Capacity Constrained Assortment Optimization, Working Paper .

## A. Appendix

Algorithm 1 gives a $(1 - \epsilon)$ approximation for (10).

**Theorem 3.** *Let* $d(Q) \leq \ell$ *and* $(\hat{\mathbf{p}}, \hat{p}_0)$ *be the solution obtained by Algorithm 1. Then* $\sum_{i=0}^{n} a_i \hat{p}_i > (1 - \epsilon)R^*$, *where* $R^*$ *is the optimal value of* (10).

*Proof.* Let $(\mathbf{p}^*, p_0^*)$ be an optimal solution to (10). Let $S = \{i \geq 1 \mid p_i^* > 0\}$. In Steps 3-6 of the algorithm we consider all the solutions that have strictly less than $\lfloor \frac{\ell}{\epsilon} \rfloor$. Hence, without loss of generality assume that $|S| \geq \lfloor \frac{\ell}{\epsilon} \rfloor$ and $S = \{1, 2, \ldots, k\}$ for some $k \geq \lfloor \frac{\ell}{\epsilon} \rfloor$ and $a_k \leq a_{k-1} \leq \cdots \leq a_1$. Also, let $S_1 = \{1, 2, \ldots, k^*\}$ where $k^* = \lfloor \frac{\ell}{\epsilon} \rfloor$. Note that $p_1^* = p_2^* = \cdots = p_k^* = p_0^*$. Therefore,

$$a_k p_k^* \leq a_{k-1} p_{k-1}^* \leq \cdots \leq a_1 p_1^*,$$

which implies

$$a_{k^*+1} p_{k^*+1}^* < \frac{\epsilon}{\ell} R^*.$$

Now consider a feasible point of (10), $(\mathbf{p}_1, p_{10})$ defined as

$$p_{10} = \frac{1}{c_0 + \sum_{i \in S_1} c_i}, \quad p_{1i} = \begin{cases} p_{10} & \text{if } i \in S_1 \\ 0 & \text{otherwise.} \end{cases}$$

implying $p_{1i}^* < p_{1i}$ for all $i \in S_1$. and since $(\mathbf{p}_1, p_{10})$ is a feasible point to (10), it follows that

$$\sum_{i \in S_1} a_i p_{1i} = \sum_{i=1}^{n} a_i p_{1i} < R^* \text{ which implies } \sum_{i \in S_1} a_i p_{1i}^* \leq R^*$$

By construction of $z_{\mathsf{LP}}(1)$, we must also have $p_{1i}^* = 0$ for every $i > k^*$ and $a_{k^*} \leq a_i$, implying

$$a_i p_{1i}^* < a_{k^*} p_{k^*}^* < \frac{\epsilon}{\ell} R^* \quad \text{for all } i > k^*.$$

Observe that $z_{\mathsf{LP}}(1) \geq R^*$ and the variables $i$ in the extreme point $(\mathbf{p}_1^*, p_{10}^*)$ that can be fractional are $i > k^*$. Therefore, $a_i p_{1i}^* < (\epsilon \cdot R^*)/\ell$ for all $i \in \mathcal{F}(\mathbf{p}_1^*, p_{10}^*)$. Thus by Lemma 1, it follows that $\sum_{i \in \mathcal{F}(p^*)} a_i p_i^*(1) < \epsilon R^*$, which implies

$$(1 - \epsilon)R^* \leq z_{\mathsf{LP}}(1) - \epsilon R^* < \sum_{i=0}^{n} a_i \hat{p}_i(1).$$

$\square$

---

**Algorithm 1** PTAS for (10)

1: Set $\mathcal{S} = \left\{ S_t \subset [n] \mid |S_t| \leq \left\lfloor \frac{\ell}{\epsilon} \right\rfloor \right\}$.

2: **for** $S_t \in \mathcal{S}$ **do**

3:      **if** $|S_t| < \left\lfloor \frac{\ell}{\epsilon} \right\rfloor$ **then**

$$\hat{p}_{t0} = \frac{1}{c_0 + \sum_{i \in S_t} c_i}$$

$$\hat{p}_{ti} = \begin{cases} \hat{p}_{t0} & \text{if } i \in S_t \\ 0 & \text{otherwise} \end{cases}$$

4:      **end if**

5:      **if** $(\hat{\mathbf{p}}_t, \hat{p}_{t0})$ feasible for (10) **then**

$$\text{Set} \quad R_t = \sum_{i=0}^{n} a_i \hat{p}_{ti}$$

6:      **else**

$$Q_t = \left\{ i \mid i \notin S_t \text{ and } \exists j \in S_t \text{ s.t. } a_j \leq a_i \right\}$$

     Consider modified (11), $z_{\mathsf{LP}}(t)$ with additional constraints

$$p_i = p_0, \ \forall i \in S_t$$
$$p_i = 0, \ \forall i \in Q_t$$

7:          **if** $z_{\mathsf{LP}}(t)$ is feasible **then**

8:          Set $(\mathbf{p}_t^*, p_{t0}^*)$: optimal extr pt of $z_{\mathsf{LP}}(t)$

9:          Set $\hat{S}_t = \left\{ i \mid p_{ti}^* = p_{t0}^* \right\}$

10:          Obtain $(\hat{\mathbf{p}}_t, \hat{p}_{t0})$ as follows:

$$\hat{p}_{t0} = \frac{1}{c_0 + \sum_{i \in S_t} c_i}$$

$$\hat{p}_{ti} = \begin{cases} \hat{p}_{t0} & \text{if } i \in \hat{S}_t \\ 0 & \text{otherwise} \end{cases}$$

11:          Set $R_t = \sum_{i=0}^{n} a_i \hat{p}_{ti}$

12:          **end if**

13:      **end if**

14: **end for**

15: Set $t^* = \underset{t}{\mathrm{argmax}}\ R_t$;

16: Output $(\hat{\mathbf{p}}, \hat{p}_0) = (\hat{\mathbf{p}}_{t^*}, \hat{p}_{t^*0})$

---